UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

CRT-603

AUTOMOBILE INSURANCE RATEMAKING IN THE PRESENCE OF ASYMMETRICAL INFORMATION*

by

Georges Dionne[1]
and
Charles Vanasse[2]

[1] Département de Sciences économiques and Centre de recherche sur les transports, Université de Montréal.

[2] Centre de recherche sur les transports, Université de Montréal.

* Preliminary version for private discussion only. Comments are welcome. This research was financed by the Social Sciences and Humanities Research Council of Canada and was presented at a seminar at the Center for Research on Risk and Insurance, University of Pennsylvania.

Université de Montréal
Centre de recherche sur les transports - Publication #603
Département de Sciences économiques - Cahier #8834
November 1988

Dépôt légal - 4e trimestre 1988

ABSTRACT

Automobile insurance is an example of a market where multiperiod contracts are observed. This form of contract can be justified by asymmetrical information between the insurer and the insured. Since the observed length of contracts may not be long enough to permit statistical convergence, insurers also use risk classification together with bonus-malus systems. In this paper, we show that the actual methodology for the integration of these two approaches can lead to inconsistencies. We then develop a statistical model that integrates adequately risk classification and experience rating. For this purpose, we present Poisson and negative binomial models with regression component in order to use all available information in the estimation of accident distribution. A bonus-malus system which integrates a priori and a posteriori information on an individual basis is proposed and multivariate insurance premium tables are derived as a function of time, past accidents and the significant variables in the regression. The statistical results were obtained from a sample of 19,013 drivers.

Key words : asymmetrical information, multiperiod contracts, risk classification, bonus-malus, Poisson, negative binomial, premium tables.

- - - - - -

RÉSUMÉ

L'assurance automobile est un exemple de marché où des contrats à plusieurs périodes sont observés. Cette forme de contrats est une conséquence de la présence d'asymétrie d'information entre les assureurs et les assurés. Puisqu'en pratique la durée des contrats n'est pas suffisamment longue pour permettre la convergence statistique de la tarification basée seulement sur l'expérience passée, les assureurs utilisent aussi la classification des risques selon certaines caractéristiques des individus.

Nous montrons que la méthodologie utilisée pour intégrer les deux approches peut être incohérente. Nous proposons un modèle statistique qui intègre adéquatement les deux approches. Nous présentons des modèles de Poisson et de binomiale négative avec composantes de régression afin d'utiliser toute l'information disponible lors de l'estimation de la distribution des accidents. Nous proposons un système bonus-malus qui tient compte de l'information a priori et a posteriori sur les individus. Nous dérivons ensuite des tables de primes d'assurances en fonction des périodes, des accidents passés et des variables significatives des régressions. Nos résultats statistiques sont obtenus à partir d'un échantillon de 19 013 conducteurs du Québec.

Mots-clés : asymétrie d'information, contrats à plusieurs périodes, classification des risques, bonus-malus, Poisson, binomiale négative, tables de primes.

1. Introduction

Since the early 1980's, economists have been very active in the study of multiperiod contracts. There are at least two reasons for the recent interest in this type of contracts. First, single-period contracts do not always produce satisfactory theoretical results in the presence of asymmetrical information and, second, multiperiod contracts are often observed in many markets or sectors of economic activity. One type of research on optimal contracts studies conditions under which multiperiod contracts dominate single-period contracts (Rubinstein and Yaari 1983, Radner 1981,1985; Dionne and Lasserre 1985,1987) while another type of research studies the characteristics of observed contracts in different sectors of the economy. An example of such a study is found in Ruser (1985) which explains why there is a smaller relationship between worker's compensation benefits and injury rates in larger firms.

Automobile insurance is another example of a market where multiperiod contracts are used. In many European countries, in Japan and in some North-American states or provinces, public and private insurers use bonus-malus systems (or experience rating) in order to relate insurance premiums to the individual's past experience (Lemaire, 1985, 1988). Insurers also base risk classification on variables which are easy to observed such as age, sex, type of driver's licence, type of car, etc... These two forms of ratemaking can be justified by asymmetrical information between the insurer and the insured[1]. The insurer does not know, ex-ante, the parameters of the true individual's distribution of accidents (i.e. adverse selection) and does not observe, during the length of the contract, the individual's actions or self-protection activities (i.e. moral hazard).

Many methodologies can be used to estimate the individual's expected number of accidents. In this paper, we are concerned with the following methodology which is often presented in the literature (Lemaire 1985, Eeghen et al. 1983). In a first stage, linear regression models are

estimated in order to a) identify significant risk classification variables; b) determine tariff classes; and, c) calculate premiums (a priori model). Usually, these classes are not homogeneous and they may generate a ratemaking structure that is unfair to the insured driver. Moreover, risk classification does not take into account moral hazard. In order to both reduce the gap between the individual's premium and risk and to increase incentives for road safety, in the second stage, the individual's past record is taken into consideration under a bonus-malus system (a posteriori model). Using Bayes' Theorem, the individual's premium is calculated as a function of time and previous accidents.

The objective of this paper is to develop a statistical model that integrates the two stages of the analysis into a single model. The model also generalizes the above methodology. We first show that, when the two stages of the analysis are not adequately integrated, inconsistencies are produced. The Poisson and the negative binomial models with a regression component are then introduced and estimated in order to use all available information in the estimation of accident distribution. It is shown that the negative binomial model approximates reasonably the true accident distribution. A bonus-malus system which integrates a priori and a posteriori information on an individual basis is then proposed and multivariate insurance premium tables are derived as a function of time, past accidents and the significant variables in the regression component. The statistical results were obtained from a sample of 19,013 drivers.

2. Risk Classification and Bonus-Malus Systems : Review of the Literature

2.1. Economic Reasons for Risk Classification and Experience Rating

Let us consider, for a moment, an insurer who has to evaluate, ex-ante, the risk of a potential client and let us limit ourselves to actuarial insurance premiums (i.e. there are no administrative costs).

In the presence of full information, the insurer knows the individual's expected number of accidents (the severity of accidents is not considered in this study) and can use this information for the purpose of setting premiums. It is widely accepted that, under these circumstances, the insurer will offer full insurance under fair conditions to any type of risk.

Many authors have studied the effect of asymmetrical information on insurance markets. One major finding which concerns us here is the fact that risk classification is efficient when it is based on variables observable at a low cost (Hoy 1982, Crocker and Snow, 1986). However, the analysis assumes that the insured individual cannot affect his distribution of accidents by engaging in self-protection activities. In other words, they limit themselves to the adverse selection problem.

Other contributions have shown that partial insurance coverage limits the moral hazard problem in single-period contracts. More recent studies presented conditions under which multiperiod insurance contracts increase the welfare of the insurer and the insured under moral hazard (Radner 1981, 1985; Rubinstein and Yaari 1983). Dionne and Lasserre (1985) obtained similar results in a model with adverse selection but, more importantly, they have shown that the same results hold when both problems are investigated simultaneously (Dionne and Lasserre, 1987). Moreover, their analysis demonstrates that some synergistic effects might occur by combining moral hazard and adverse selection problems in models which use past experience. Indeed, the same information required to reduce either moral hazard alone or adverse selection alone, can be used to reduce the effects of both problems simultaneously. For example, past experience provides information on the individual's distribution of accidents, which is a function of both driving ability and self-protection activities.

In their analysis of the foundation of bonus-malus systems, Henriet and Rochet (1986) have isolated the basic functions of bonus-malus systems. In the presence of adverse selection, the focus has to be on the estimation of the individual's expected number of accidents while, in

the presence of moral hazard, it has to be on constant safety incentives. Consequently, in the first case, observations carried out over many periods are statistically useful while, in the second case, recent past experience should have more weight in the overall past record.

Insurers use risk classification and bonus-malus systems (Lemaire, 1985). With few exceptions, insurers do not limit themselves to bonus-malus systems for obvious reasons of statistical convergence and they do not restrict their activities to simple risk classification since the latter does not produce risk classes sufficiently homogeneous and does not take into account road safety incentives. In other words, bonus-malus systems have two functions : first, they reduce statistical dispersion within risk classes and, second, they motivate constant road safety activities.

2.2 Empirical Studies of the Automobile Insurance Market

Few researchers have explicitly investigated asymmetrical information problems in empirical studies on automobile insurance. Among them, Dahlby (1983) has studied the discriminatory effect of ratings based on sex, D'Arcy and Doherty (1988) have investigated multiperiod price stategies in presence of adverse selection and Boyer and Dionne (1989) have verified the empirical proposition that past experience is a good predictor of the individual's risk. They also constructed a multivariate, multiperiod pricing formula for automobile insurance in the spirit of Radner-Rubinstein-Yaari theoretical contract. However, Boyer and Dionne (1989) did not investigate bayesian models and did not calculate insurance premiums. Our article focuses on these issues.

3. Presentation and Criticism of the Current Bayesian Methodology

Lemaire (1985, 1988) discusses in detail bayesian bonus-malus systems and presents empirical estimations of such these systems. He

also investigates risk classification based on the driver's characteristics (see also Van Eeghen, Greup and Nijssen 1983, Ferreira 1974, Seal 1969 and Greenwood and Yule 1920).

The methodology is divided into two main stages of analysis : a priori and a posteriori tarification. A priori tarification refers to the use of significant variables from ordinary least-squares regression models in order to determine risk classes and to estimate the corresponding expected number of accidents. Then, for reasons mentioned previously, bayesian analysis allows for the adjustment of premiums.

In a posteriori tarification, univariate Poisson and negative binomial models are considered. The negative binomial model usually produces a more accurate approximation of true accident distribution. Finally, the derived bayesian bonus-malus system is compatible with the financial and actuarial objectives of the insurers. Nevertheless, many improvements of the methodology are needed.

A first criticism concerns the use of ordinary least-squares econometric models for identification of rating variables. Research conducted over many years has shown that automobile accidents follow a discrete stochastic process (of the Poisson type) characterized by many "zero accident" observations and by the absence of negative observations. Ordinary least squares estimators are not appropriate here since they allow for negative predicted expected number of accidents. Moreover, by definition, ordinary least squares imply that the marginal effect of each explanatory variable is constant and independent of other variables. Finally, individual accident data may cause heteroskedasticity and usual methods of correction could affect the results (see Lemaire (1985) for more detailed comments on linear models).

A second criticism is linked to the fact that univariate (without regression component) statistical models are used in the bayesian determination of the individual's insurance premiums. Consequently, insurance premiums are function merely of time and of the past number of accidents. The premiums do not vary simultaneously with other variables that affect accident distribution. The most interesting example is the

age variable. Let us suppose, for a moment, that age has a significant negative effect on the expected number of accidents. This implies that insurance premiums should decrease with age. Premium tables derived from univariate models do not allow for a variation of age, even if they are a function of time. However, a general model with a regression component would be able to determine the specific effect of age when the variable is statistically significant.

Finally, the third criticism concerns the coherency of the two-stage procedure using different models in order to estimate the same distribution of accidents (linear and nonlinear models).

In the following section we will introduce a methodology which responds adequately to the three criticisms. First, count data models will be proposed to estimate the individual's accident distribution. The main advantage of the count data models over the linear regression model lies in the fact that the dependent variable is a count variable restricted to non-negative values. Statistical results from Poisson and negative binomial distributions with a regression component are presented in the next section of the paper[2]. Second, we will generalize the optimal bonus-malus system by introducing all information from the regression into the calculation of premium tables. These tables will take into account time, accident record and the individual characteristics.

4. Estimation of Poisson and Negative Binomial Models with a Regression Component

4.1. The Models

It is now well accepted that the Poisson distribution does satisfy the basic characteristics associated to the estimation of the probability that a individual i is involved in Y_i automobile accidents during a given period : accidents are independent and the probability of having more than one accident in a given period is small. Assuming the Poisson

distribution, the probability that individual i will be involved in Y_i accidents during a given period is :

$$Pr(Y_i) = \frac{e^{-\exp(x_i\beta)}(\exp(x_i\beta))^{Y_i}}{Y_i!}, \quad i = 1,\ldots,I \qquad (1)$$

where $x_i(= (x_{i1}, x_{i2},\ldots,x_{ik}))$ is a vector of individual's characteristics, and β is a vector of coefficients.

For notational convenience, let us denote $\lambda_i = (\exp(x_i\beta)$. It is well known that $\lambda_i = E(Y_i) = Var(Y_i)$. This rules out overdispersion. We should emphasize that λ_i is not a random variable; the model assumes implicitly that the k exogenous variables provide enough information to obtain the appropriate estimation of the individual's probability. The randomness is due to the Poisson specification. When the regression component is a constant c, $E(Y_i) = Var(Y_i) = \exp(c)$, which implies that all individuals are considered identical as in the univariate model.

If we now assume that the vector of explanatory variables does not contain all the significant information, a random variable has to be introduced into the regression component. Following Gourieroux, Monfort and Trognon (1984), λ_i can be written as

$$\lambda_i = \exp(x_i\beta + \epsilon_i)$$

or equivalently as

$$\lambda_i = \exp(x_i\beta)u_i$$

where $u_i \equiv \exp(\epsilon_i)$.

It is convenient to suppose that u_i follows a gamma distribution with parameters (a,a), implying that $E(u_i) = 1$ and $Var(u_i) = 1/a$. We will show later that this parameterization of the gamma distribution does not affect the results if there is a constant in the regression component[3]. The probability that individual i will be involved in Y_i accidents becomes

$$Pr(Y_i) = \int_0^{\infty} Pr(Y_i/u_i) f(u_i)\, du_i$$

$$= \frac{\Gamma(Y_i+a)}{Y_i!\,\Gamma(a)} \left[\frac{\exp(x_i\beta)}{a}\right]^{Y_i} \left[1 + \frac{\exp(x_i\beta)}{a}\right]^{-(Y_i+a)} \qquad (2)$$

which is a negative binomial distribution with parameters $(\exp(x_i\beta), a)$. Then $E(Y_i) = \exp(x_i\beta)$ and $Var(Y_i) = \exp(x_i\beta)\left[1 + \frac{\exp(x_i\beta)}{a}\right]$. We observe that $Var(Y_i)$ is a nonlinear increasing function of $E(Y_i)$, so the model allows overdispersion. When the regression component is a constant c, $E(Y_i) = \exp(c)$ and $Var(Y_i) = \left[1 + \frac{\exp(c)}{a}\right]$ which correspond, respectively, to the mean and variance of the univariate negative binomial distribution with parameters $(a, 1/b)$ where $b = \exp(c)/a$. In section 4.3, maximum likelihood estimation results for the Poisson and negative binomial regressions will be presented.

There is some arbitrariness in the choice of $f(u_i)$ when deriving the negative binomial model. In some respects, it is important to explore the consequences of this choice. If the "true" distribution of u_i is not a gamma, the maximum likelihood estimators (MLE) obtained from the negative binomial model will not be consistent. Consequently, it would be inappropriate to use the MLE in a tarification model based on the negative binomial distribution. Gourieroux, Monfort and Trognon (1984) showed that, if the distribution of accidents belongs to the linear exponential family (as in the case of the Poisson and negative binomial distributions) and if the mean is correctly specified, "pseudo" MLE (PMLE) estimators will be consistent regardless of the "true" compounding distribution (see also Cameron and Trivedi (1986)). They also proposed "quasi-generalized" PMLE (QGPMLE) which are more efficient than PMLE when the mean and variance are correctly specified. Therefore, if the MLE,

PMLE and QEPMLE are very similar, one can reasonably assume that the model is well specified when assuming that u_i follows a gamma distribution. Since the tarification model presented in Section 5 is based on strong distributional assumptions, the following sections will also analyse PMLE and QGPMLE in conjunction with usual goodness of fit criteria in order to verify that the negative binomial assumption is a reasonable one.

4.2. Data and Variables

The data come from a random sample of 19 013 automobile drivers selected from the entire population of Québec automobile drivers for the year 1982-1983. The sample was built in collaboration with the Régie de l'Assurance Automobile du Québec (RAAQ), a provincial state-owned insurance company[4].

The data contain information on accidents and individual's characteristics appearing on the driver's licence : age, sex, period of validity, place of residence, driver's restrictions, class of driver's licence and experience. The sample mean and sample variance of the dependent variable are respectively 0.0710 and 0.07757 indicating some overdispersion. Table 1 describes the variables used in the econometric analysis.

Insert Table 1 here.

4.3. Estimation Results

For the sake of comparison with traditional models, we have estimated the parameters of both the univariate Poisson and the univariate negative binomial models by the maximum likelihood method. Expected frequencies were generated and were compared with the observed frequencies. Results are presented in Table 2. The χ^2 test indicates that the Poisson distribution is rejected at the 5% significance level

while the negative binomial model is not rejected. These results indicate that a better fit is obtained when some heterogeneity among the insured individuals is assumed.

Insert Table 2 here.

As discussed in the previous section, the Poisson model with a regression component takes into account some differences among drivers. Therefore, despite the previous results, this model was estimated along with the negative binomial model with a regression component[5]. The results are summarized in Table 3. They show that age and sex interaction variables are significant (Female 19-year olds or younger is the omitted category) as well as classes of driver's licences for bus, truck and taxi drivers. As suggested in another study (Boyer and Dionne, 1989) these variables are proxies for risk exposure. It is interesting to observe that only one place of residence (Outaouais) is statistically different from Montréal (omitted place of residence). This last result challenges the often observed practice by the industry of using the place of residence for insurance tarification.

Even if the Poisson model gives results similar to the negative binomial model, it is important to test whether there is a gain in efficiency by using a model allowing for overdispersion of the data. Since both models are estimated by the maximum likelihood method and since the Poisson model can be derived from the negative binomial model by a single parametric restriction (on a), one can apply a standard likelihood ratio test. Under the null hypothesis (H_o), the test statistic is 25.98 (-2(-4661.57 + 4648.58)). Under H_o, this statistic is asymptotically distributed as a χ_1^2, so we reject the null hypothesis (or the Poisson model) since the critical value of this distribution at a 5 per cent significance level is 3.84.

It is also important to test if there is a gain when estimating a negative binomial model with a regression component over a standard

univariate negative binomial model. As pointed out above the univariate model is a limiting case of the general model with a regression component.

One can test the significance of the regression component by a likelihood ratio test since both models are nested (setting all β coefficients, except the constant term, to zero gives the univariate model). Under the null hypothesis, the test statistic is 536.4 (-2(-4916.78 + 4648.58)) and is asymptotically distributed as a χ_{52}^2. We reject the univariate model (the critical value of the test at a 5 per cent significance level is 69.83).

From those tests we verified that the negative binomial model with a regression component performed better than both the Poisson and the univariate negative binomial models but this still leaves open the question of whether the negative binomial assumption is a reasonable assumption. In order to verify this assumption, the model has been estimated by PML and QGPML methods. We obtained, with both methods[6], results very similar to the MLE. We can then reasonably assume that the data follow a negative binomial distribution.

Insert Tables 3 and 4 here.

5. Optimal Bonus-Malus System

The results of the preceding section indicate that the negative binomial model with a regression component can be applied to insurance ratemaking. Since many characteristic variables are significant, a priori and a posteriori information can be integrated in a unified and general model. In this section of the article, we present a generalization of the bayesian bonus-malus system in order to integrate the a priori information based on individual's characteristic. Then we apply the statistical results obtained in Section 4 to the generalized model.

5.1. A General Bonus Malus System

Consider again an insured driver i with an experience over t periods; let Y_i^j represent the number of accidents in period j and x_i^j, the vector of the k characteristics observed at period j, that is $x_i^j = (x_{i1}^j, \ldots\ldots, x_{ik}^j)$. Let us further suppose that the true expected number of accidents of individual i at period j, $\lambda_i^j(u_i, x_i^j)$, is a function of both individual characteristics x_i^j and a random variable u_i. The insurer needs to calculate the best estimator of the true expected number of accidents at period t+1. Let $\bar{\lambda}_i^{t+1}(Y_i^1, \ldots\ldots, Y_i^t; x_i^1, \ldots\ldots, x_i^{t+1})$ designate this estimator which is a function of past experience over the t periods and of known characteristics over the t+1 periods.

One can show that, under conditions similar to those presented by Lemaire (1985), the optimal estimator is equal to:

$$\bar{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^{t+1}) = \int_0^\infty \lambda_i^{t+1}(u_i, x_i^{t+1})\, f(\lambda_i^{t+1}/Y_i^1, \ldots, Y_i^t; x_i^1, \ldots\ldots, x_i^t) d\lambda_i^{t+1} \quad . \tag{3}$$

where, by the Bayes' rule,

$$f(\lambda_i^{t+1}/Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^t) = \frac{P((Y_i^1, \ldots, Y_i^t)/\lambda_i^{t+1}; x_i^1, \ldots\ldots, x_i^t) f(\lambda_i^{t+1})}{\bar{P}((Y_i^1, \ldots\ldots, Y_i^t)/x_i^1, \ldots\ldots, x_i^t)} \tag{4}$$

and where, by definition,

$$\bar{P}((Y_i^1, \ldots, Y_i^t)/x_i^1, \ldots, x_i^t) = \int_0^\infty P((Y_i^1, \ldots, Y_i^t)/\lambda_i^{t+1}; x_i^1, \ldots, x_i^t)\, f(\lambda_i^{t+1}) d\lambda_i^{t+1} \tag{5}$$

Applying the negative binomial distribution to the model, the probability of the sequence $(Y_i^1, \ldots\ldots, Y_i^t)$ given the true expected number of accidents at t+1 and the individual's characteristics, is a t-dimension Poisson distribution

$$P((Y_i^1,\ldots,Y_i^t)/\lambda_i^{t+1};\ x_i^1,\ldots,x_i^t) = \frac{e^{-\sum_{j=1}^{t}\lambda_i^j}\ \prod_{j=1}^{t}(\lambda_i^j)^{Y_i^j}}{\prod_{j=1}^{t}(Y_i^j!)} \qquad (6)$$

where $\lambda_i^j = \exp(x_i^j\beta)u_i \equiv (\dot{\lambda}_i^j)u_i$. The unconditional distribution of $(Y_i^1,\ldots,Y_i^t)$ when u_i follows a gamma distribution with mean equal to one and variance equal to 1/a is given by

$$\bar{P}\ (Y_i^1,\ldots,Y_i^t/x_i^1,\ldots,x_i^t) = \frac{\Gamma(a+\bar{Y}_i)\ a^a\ \prod_{j=1}^{t}(\dot{\lambda}_i^j)^{Y_i^j}}{\prod_{j=1}^{t}(\lambda_i^j!)\ \Gamma(a)\ (a+\bar{\lambda}_i)^{\bar{Y}_i+a}} \qquad (7)$$

where $\bar{Y}_i = \sum_{j=1}^{t} Y_i^j$ and $\bar{\lambda}_i = \sum_{j=1}^{t}\dot{\lambda}_i^j$.

Using Bayes' rule, one verifies that

$$f(\lambda_i^{t+1}/Y_i^1,\ldots,Y_i^t;x_i^1,\ldots,x_i^t) = \frac{[a+\bar{\lambda}_i]^{\bar{Y}_i+a}\ e^{-u_i(a+\bar{\lambda}_i)}\ u_i^{a-1+\bar{Y}_i}}{\Gamma(a+\bar{Y}_i)} \qquad (8)$$

which corresponds to a gamma distribution with parameters $(a+\bar{Y}_i,\ a+\bar{\lambda}_i)$.

Therefore, the optimal estimator of the true expected number of accidents for individual i is:

$$\hat{\lambda}_i^{t+1}(Y_i^1,\ldots,Y_i^t;x_i^1,\ldots,x_i^{t+1}) = \dot{\lambda}_i^{t+1}\left[\frac{a+\bar{Y}_i}{a+\bar{\lambda}_i}\right] \qquad (9)$$

When t=0, $\hat{\lambda}_i^1=\dot{\lambda}_i^1 \equiv \exp(x_i^1\beta)$ which implies that only a priori tarification is significant in the first period. Moreover, when the regression component is limited to a constant c, one obtains:

$$\tilde{\lambda}_i^{t+1}(Y_i^1,\ldots,Y_i^t) = \exp(c)\left[\frac{a+\bar{Y}_i}{a+t\,\exp(c)}\right] = \frac{a+\bar{Y}_i}{(1/b)+t} \qquad (10)$$

which corresponds to the well known univariate model. This result is not affected by the parameterization of the gamma distribution.

Two criteria define an optimal bonus-malus system which has to be fair for the policyholders and be financially balanced for the insurer. It is clear that the estimator proposed in (9) is fair since it allows the estimation of the individual's risk as a function of both his characteristics and past experience. Dionne and Vanasse (1988) shows that the above model is also financially balanced.

5.2. Application of the General Bonus-Malus System

For purposes of comparison, we begin the application by calculating a table of premiums derived from the univariate negative binomial model or, equivalently, from a regression component limited to a constant c.

If we assume that the first premium is equal to $100, the premium in the period (t+1) is then equal to

$$\tilde{P}_i^{t+1}(Y_i^1,\ldots,Y_i^t) = 100\,\frac{1/\tilde{b}\;(\hat{a}+\bar{Y}_i)}{\hat{a}\;(1/\tilde{b}+t)} \qquad (11)$$

where $\hat{a}$ and $\tilde{b}$ are the estimated parameters. Using the values presented in Table 2, the following table can be constructed:

Insert Table 5 here.

We observe that only two variables may change the level of insurance premiums that is time (t) and the number of accumulated accidents ($\bar{Y}_i$). For a given number of accidents, as t increases the expected number of accidents decreases and the corresponding premiums decrease as well. We also observe that the premiums increase quickly when the number of accumulated accidents increases. For example, an insured who had two

accidents in the first period will pay a premium of $322.42 in the next period, but if he had no accidents, he would have paid only $90.86.

From (10), it is clear that no a priori information can be introduced in order to differentiate an individual's risk. However, from (9), a more general pricing formula can be derived:

$$\hat{P}^{t+1}(Y_i^1,\ldots,Y_i^t;x_i^1,\ldots,x_i^{t+1}) = M\,\tilde{\lambda}_i^{t+1}\left[\frac{\hat{a}+\bar{Y}_i}{\hat{a}+\tilde{\bar{\lambda}}_i}\right] \qquad (12)$$

where $\tilde{\lambda}_i^{t+1} \equiv \exp(x_i^{t+1}\,\tilde{\beta})$, $\tilde{\bar{\lambda}} \equiv \sum_{j=1}^{t} \exp(x_i^j\,\tilde{\beta})$, and M is such that

$1/I \sum_{j=1}^{I} \tilde{\lambda}_i^{t+1}\, M = \$100$ when the total number of insureds is I.

This general pricing formula is a function of time, the number of accumulated accidents and the individual's significant characteristics in the regression component. Tables 6 and 7 are contructed from (12). Table 6 is for a male who had his first driving licence at 30 years old, who lives in Montreal and holds the standard class of driver's licence (42). Many differences can be observed between Table 5 and Table 6. First, contrarily to the univariate model, the driver in Table 6 does not start with a premium of $100. Second, since the age variable is significant in the estimated model, two factors rather than one have a negative effect on the individual's premiums (i.e. time and age). In Table 6, the premium is largely reduced when the driver reaches 35 years of age in period 5.

Insert Table 6 here.

Table 7 presents the case of a female. The only differentiating characteristic with the example of Table 6 is gender. The fact of being a female reduces significantly the premiums.

Insert Table 7 here.

Finally, when the insured modifies significant variables, new tables may be constructed. Table 8 is for the individual in Table 6 who decided to become a Montreal taxi driver (class 31) in period 6. Having two accidents, he now pays $377.42 instead of $148.37.

6. Conclusion

In this paper we have presented a statistical model that integrates a priori and a posteriori tarification models into a single model. We have shown how a bonus-malus system can be modified in order to take into account a priori information on an individual basis. The proposed model permits to eliminate inconsistencies associated with the two-stage procedure.

We have also shown that, based on the sample used, the negative binomial model with a regression component produces a reasonable approximation of the true distribution of accidents. Different statistical tests have rejected the Poisson model and have shown that the negative binomial model with a regression component performs better than the univariate negative binomial model. PML and QGPML methods of estimation provided very similar results to those obtained from the negative binomial model with a regression component. Finally, we have extended the standard bayesian tarification model and have shown how the empirical results can be utilized for the calculation of multivariate tables of insurance premiums.

A possible line of future research is the integration of accident severity into the general model. Recent contributions have analysed different types of distribution functions to be applied to the severity

of losses (Lemaire 1985 for automobile accidents, Cummins et al. 1988 for fire losses, and Hogg and Klugman 1984, for many other applications). To our knowledge, no study has ever considered the possibility of introducing the individual's characteristics or action in a model that simultaneously explains the occurrence and the severity of accidents. This investigation would enable insurers to build a fair rating system and to increase incentives for safety under asymmetrical information.

References

Boyer, M. and G. Dionne, "An Empirical Analysis of Moral Hazard and Experience Rating" Review of Economics and Statistics (1989 forthcoming).

Boyer, M., G. Dionne and R. Kihlstrom, "Public Information and Experience Rating" Mimeo, Université de Montréal, 1986. Forthcoming in Studies in the Economics of Uncertainty : in Honor of Josef Hadar, (New-York : Springer-Verlag).

Cameron A.C. and R.K. Trivedi, "Econometric Models Based on Count Data: Comparison and Applications of Some Estimators and Tests" Journal of Applied Econometrics 1 (1986), 29-53.

Crocker, K.J. and A. Snow, "The Efficiency Effects of Categorical Discrimination in the Insurance Industry" Journal of Political Economy, (1986).

Cummins J.D., G, Dionne, J.B. McDonald and B.M. Pritchett "Application of the GB2 Distribution in Modeling Insurance Loss Processes" Mimeo, Department of Insurance, University of Pennsylvania (1988).

Dahlby, B.A., "Adverse Selection and Statistical Discrimination" Journal of Public Economics 20 (1983), 121-131.

D'Arcy, S. and N. Doherty, "Adverse Selection, Private Information, and Low Balling in Insurance Markets" mimeo, University of Pennsylvania (1988).

Dionne G. and P. Lasserre, "Adverse Selection, Repeated Insurance Contracts and Announcement Strategy" Review of Economic Studies, 70 (1985), 719-723.

Dionne G. and P. Lasserre, "Dealing with Moral Hazard and Adverse Selection Simultaneously" Working paper, Center for the Study of Risk and Insurance, University of Pennsylvania, (1987).

Dionne G. and C. Vanasse "A Generalization of Actuarial Automobile Insurance Rating Models" Mimeo, CRT and Economics Department, Université de Montréal (1988).

Ferreira, J., "The Long-Term Effects of Merit-Rating Plans on Individual Motorists" Operations Research 22 (1974), 954-978.

Gourieroux C, A. Monfort and A. Trognon "Pseudo Maximum Likehood Methods: Application to Poisson Models" Econometrica, 52 (1984), 701-720.

Greenwood M. and G.U. Yule "An Inquiry into the Nature of Frequency Distribution of Multiple Happenings "Journal of Royal Statistical Society A, 83, (1920), 255-279 .

Hall, B.H., A. Griliches and J.A. Hausman, "Patents and R and D : Is here a Lag" International Economic Review 27 (1986), 265-283.

Hausman J.A. B.H. Hall and Z. Griliches "Econometric Models for Count Data with an Application to the Patents - R & D Relationship" Econometrica, 52 (1984), 910-938.

Henriet D. and J.C. Rochet "La logique des systèmes bonus-malus en assurance automobile : une approche théorique" Annales d'Economie et de Statistique, (1986), pp 133-152.

Hoy, M., "Categorizing Risks in the Insurance Industry" Quarterly Journal of Economics, 97 (1982), 321-337.

Hogg R.V. and S.A. Klugman Loss Distributions. (New-York: Wiley, 1984).

Lemaire J. Automobile Insurance: Actuarial Models. (Boston: Kluwer-Nighoff, 1985)

Lemaire J. "A Comparative Analysis of Most European and Japanese Bonus-Malus Systems" Mimeo, Insurance Department, University of Pennsylvania (1988).

Palfrey, T.R. and C.S. Spatt "Repeated Insurance Contracts and Learning" Rand Journal of Economics 16(1985), 356-367.

Radner, R. "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship" Econometrica 49(1981), 1127-1148.

Radner, R. "Repeated Principal-Agent Games with Discounting" Econometrica 53 (1985), 1173-1199.

Rubinstein, A. and M.E. Yaari, "Repeated Insurance Contracts and Moral Hazard" Journal of Economic Theory 30 (1983), 79-97.

Rose, N. "Profitability and Product Quality : Financial Indicators and Airline Safety Performance" Mimeo, M.I.T. (1988).

Ruser, J.H. "Workers' Compensation Insurance, Experience Rating, and Occupational Injuries" The Rand Journal of Economics, 16 (1985), 487-503.

Seal H.L. (1969) Stochastic Theory of a Risk Business. (New-York: Wiley, 1969).

Van Eeghen J., E.K., Greup and J.A., Nijssen Surveys of Actuarial Studies: Rate Making. Research Department, Nationale - Nederlanden N.V.

NOTES

[1]Another complementary explanation is learning. See Boyer, Dionne and Kihlstrom (1986) and Ralfrey and Spatt (1985).

[2]See N. Rose for an application of the Poisson model in the airline market.

[3]For other parametrizations of the gamma distribution, see Cameron and Trivedi (1986).

[4]See Boyer and Dionne 1989, for more details on the data and on the Québec automobile insurance plan.

[5]See Hausman, Hall and Griliches 1984, Hall, Griliches and Hausman 1986, Gourieroux, Monfort and Trognon 1984 and Cameron and Trivedi 1986 for detailed presentations of these econometric models.

[6]Only QGPMLE are presented in Table 4. Those with the PML are available upon request.

Table 1

List of variables used in the econometric analysis

AGE - SEX: 14 dichotomous variables

F 19 = [1 if, on August 1 1982, the driver is a female under 19 years of age
0 otherwise (reference group).

M 2024 = [1 if, on August 1 1982, the driver is a male between 20 and 24.
0 otherwise

etc ...

EXPERIENCE: 7 dichotomous variables to measure the number of years since a driver obtained a first permit

EXP0 = 1 if permit obtained in 1982-1983, 0 otherwise

EXPM1 = 1 if, on August 1 1982, the permit has been held for less than a year, 0 otherwise

EXP12 = 1 if permit was obtained between 1 and 2 years ago (before 1/8/82)
0 otherwise (reference group)

etc ...

PLACE OF RESIDENCE: 11 dichotomous variables to capture the effect of the immediate environment of the individual driver

REG1 = [1 if the driver lives in the Lower St.Lawrence or Gaspé Regions
0 otherwise

REG2 = Saguenay-Lac St.Jean

REG3 = Québec city region

REG4 = Trois-Rivières

REG5 = Eastern Townships

REG6 = Montreal region: reference group

REG7 = Laurentians-Lanaudière

REG8 = Montérégie

REG9 = Outaouais

REG10 = Abitibi-Témiscamingue

REG11 = North-Shore and Nouveau-Québec

DRIVER'S RESTRICTIONS: 13 dichotomous variables to capture the effect of driving restrictions appearing on the driver licence (type of license)

RTSA = 1 if the driver must wear glasses, 0 otherwise

RTSB = can drive during daylight only

RTSCG = must wear a hearing aid (C) or must drive a car with power brakes (L,N,P) or must drive a car with a left acceleration (r) or is subject to medical restrictions (S) or other restrictions (T) or has a temporary permit (W).

RTSD = must get a medical exam regularly

RTSH = can only drive a vehicle whose total weight is less than 5,500 kg.

RTSJ = must drive an automatic transmission equipped car

RTSK = must drive a car with power steering

RTSM = cannot drive a business vehicle, a bus, a minibus or a taxi

RTSO = cannot drive an emergency vehicle

RTSQ = must drive a car equipped with manual signals or a magic eye

RTSU = has a license valid for 6 months only

RTSY = cannot drive a taxi or an ambulance

RTSo = no restrictions

CLASS OF DRIVER'S LICENSE: 9 dichotomous variables to isolate the effect of the class of driving license (since class 41 and class 42 are almost mutually exclusive only class 42 is used as a variable; since no driver in our sample has class 61 only, this class is not used as a variable).

CL1112 = [1 if driver can drive a bus; include classes 13,41,42,61
         [0 otherwise

CL13 = minibus; includes 41, 42 and 61

CL21 = vehicle or set of vehicles whose weight may exceed 11000 kg; includes 22, 42 and 61

CL22 = vehicle whose weight may exceed 11000 kg; includes 41, 42 and 61

CL31 = taxis; includes 41, 42 and 61

CL41 = vehicles or set of vehicles whose weight is less than 11000 kg (includes 42 and 61)

CL42 = vehicles of class 41 except public vehicles; includes 61

CL54 = motorcycle; includes 55, 56 and 61

CL55 = motorcycle of less than 400 cm^3; includes 56 and 61

CL56 = velomotor, includes 61

CL61 = tractor

VALC = indice of validity of the driver's licence during 1982-83 (number of days where the driver's permit is valid divided by 365)

Table 2

Estimates of Univariate Poisson and Negative Binomial Distributions

Individual's number of accidents in a given period	Observed numbers of individuals during 1982-1983	Predicted numbers of individuals for 1982-1983	
		Poisson ($\exp \hat{c} = 0.0701$)	Negative binomial ($\hat{a}=0.6960; 1/\hat{b}=9.9359$)
0	17,784	17,727	17,785
1	1,139	1,242	1,132
2	79	43	88
3	9	1	7
4	2	0	1
5+	0	0	0
	19,013	$\chi^2 = 133$	$\chi^2 = 2.21$
		$\chi^2_{2,95} = 5.99$	$\chi^2_{1,95} = 3.84$
		LLF = -4950.28	LLF = -4916.78

Table 3

MLE of the Poisson and Negative Binomial Models

	NEGATIVE BINOMIAL		POISSON	
VARIABLE	COEFFICIENT	T-RATIO	COEFFICIENT	T-RATIO
CTE	-3.956	-9.132	-3.927	-10.549
M19M	1.067	4.606	1.065	4.812
M2024	1.049	4.175	1.046	4.415
M2534	.465	1.780	.464	1.877
M3544	.294	1.065	.297	1.156
M4554	.165	.584	.164	.618
M5564	.005	.016	.007	.025
M65	-.180	-.573	-.189	-.615
F2024	-.097	-.367	-.098	-.380
F2534	-.350	-1.303	-.350	-1.367
F3544	-.731	-2.461	-.731	-2.571
F4554	-.658	-2.067	-.655	-2.142
F5564	-1.382	-3.646	-1.381	-3.210
F64	-.182	-.442	-.186	-.444
EXP0	.126	.534	.104	.476
EXPM1	-.026	-.131	-.036	-.196
EXP2	.093	.516	.087	.524
EXP35	-.048	-.250	-.052	-.286
EXP610	.009	.042	-.003	-.012
EXP11	.068	.293	.057	.263
REG1	.117	.722	.117	.767
REG2	.031	.216	.025	.174
REG3	.043	.464	.037	.418
REG4	.146	1.200	.142	1.263
REG5	.001	.001	-.010	-.061
REG7	.153	1.483	.151	1.502

REG8	-.027	-.299	-.030	-.342
REG9	.409	3.066	.412	3.331
REG10	.219	1.287	.205	1.278
REG11	.182	.836	.176	.877
RTSA	-.370	-1.658	-.358	-1.858
RTSB	.272	.483	.269	.511
RTSGC	-.259	-.768	-.265	-.924
RTSD	.312	1.364	.318	1.393
RTSH	.258	.970	.271	1.325
RTSJ	1.417	1.458	1.412	1.995
RTSK	-.973	-.739	-.919	-.981
RTSM	-.626	-1.519	-.617	-1.737
RTSO	.177	.441	.185	.486
RTSQ	-1.298	-.901	-1.341	-1.136
RTSU	.588	1.655	.590	1.832
RTSY	-2.045	-1.897	-2.040	-2.019
RTSo	-.189	-.826	-.176	-.888
CL1112	.576	3.654	.556	3.356
CL13	-.588	-.521	-.573	-.571
CL21	.231	1.803	.235	1.938
CL22	.937	2.747	.943	3.395
CL31	.934	4.727	.909	4.970
CL42	-.013	-.117	-.011	-.109
CL54	.034	.244	.034	.252
CL55	-.445	-.813	-.456	-.783
CL56	-.207	-.176	-.204	-.203
VALC	1.145	4.556	1.113	5.405
1/a	.682	3.911	___	___
LLF	-4648.58		-4661.57	
Nb. of parameters	54		53	
Nb. of observations	19013		19013	

Table 4

QGPMLE of Negative Binomial Specification

VARIABLE	COEFFICIENT	T-RATIO
CTE	-3.987	-8.661
M19M	1.068	4.368
M2024	1.050	3.959
M2534	.466	1.696
M3544	.291	1.002
M4554	.166	.561
M5564	.003	.010
M65P	-.171	-.518
F2024	-.096	-.347
F2534	-.350	-1.243
F3544	-.731	-2.356
F4554	-.660	-1.987
F5564	-1.382	-3.551
F65P	-.178	-.415
EXP0	.150	.592
EXPM1	-.017	-.076
EXP2	.098	.500
EXP35	-.043	-.206
EXP610	.021	.089
EXP11	.079	.317
REG1	.119	.679
REG2	.035	.231
REG3	.051	.510
REG4	.152	1.163
REG5	.011	.061
REG7	.155	1.384

REG8	-.024	-.246
REG9	.407	2.800
REG10	.233	1.268
REG11	.189	.813
RTSA	-.382	-1.584
RTSB	.273	.457
RTSGC	-.253	-.696
RTSD	.308	1.276
RTSH	.245	.844
RTSJ	1.422	1.296
RTSK	-1.026	-.690
RTSM	-.637	-1.467
RTSO	.170	.390
RTSQ	-1.256	-.780
RTSU	.588	1.472
RTSY	-2.052	-1.851
RTSo	-.203	-.821
CL1112	.592	3.400
CL13	-.605	-.503
CL21	.226	1.630
CL22	.934	2.344
CL31	.955	4.118
CL42	-.015	-.122
CL54	.033	.217
CL55	-.437	-.739
CL56	-.213	-.165
VALC	1.179	4.487

Nb of parameters	53
Nb of observations	19,013

Table 5
Univariate Negative Binomial Model

$\hat{a}$ = .696080 $1/\hat{b}$ = 9.93580

t \ $\bar{Y}_t$	0	1	2	3	4
0	100.00				
1	90.86	221.38	351.91	462.43	612.96
2	83.24	202.83	322.42	442.01	561.60
3	76.81	187.15	297.50	407.84	518.19
4	71.30	173.72	276.15	378.58	481.00
5	66.52	162.09	257.66	353.23	448.80
6	62.35	151.92	241.49	331.06	420.63
7	58.67	142.95	227.23	311.52	395.80
8	55.40	134.98	214.56	294.15	373.73
9	52.47	127.85	203.23	278.61	353.99

Table 6

Negative Binomial Model With a Regression Component
Male, 30 years old in period 0, region 6, class 42.

t \ $\bar{Y}_t$	0	1	2	3	4
0	136.59				
1	128.23	215.62	303.01	390.40	477.79
2	120.83	203.18	285.53	367.88	450.22
3	114.24	192.09	269.95	347.81	425.67
4	108.33	182.16	255.99	329.82	403.65
5	64.73	108.84	152.96	197.07	241.19
6	62.79	105.58	148.37	191.16	233.96
7	60.96	102.51	144.05	185.60	227.15
8	59.24	99.61	139.98	180.35	220.72
9	57.61	96.87	136.13	175.39	214.65

Table 7

**Negative Binomial Model With a Regression Component
Female, 30 years old in period 0, region 6, class 42.**

t	$\bar{Y}_t$	0	1	2	3	4
0		85.84				
1		82.46	138.65	194.85	251.05	307.24
2		79.33	133.40	187.47	241.54	295.60
3		76.44	128.53	180.63	232.72	284.82
4		73.75	124.01	174.27	224.53	274.79
5		34.28	57.64	81.00	104.37	127.73
6		33.73	56.71	79.70	102.69	125.67
7		33.19	55.81	78.44	101.06	123.68
8		32.67	54.94	77.21	99.48	121.75
9		32.17	54.10	76.03	97.95	119.88

Table 8

**Negative Binomial Model With a Regression Component
Male, 30 years old in period 0, region 6, class 42 in periods 0-5 and class 31 (taxi) in periods 6-9.**

t	$\bar{Y}_t$	0	1	2	3	4
0		136.59				
1		128.23	215.62	303.01	390.40	477.79
2		120.83	203.18	285.53	367.88	450.22
3		114.24	192.09	269.95	347.81	425.67
4		108.33	182.16	255.99	329.82	403.65
5		64.73	108.84	152.96	197.07	241.19
6		159.71	268.57	377.42	486.27	595.12
7		148.40	249.54	350.68	451.81	552.95
8		138.58	233.03	327.47	421.92	516.37
9		129.98	218.56	307.15	395.73	484.32